Student Guides Teacher: Weak-to-Strong Inference via Spectral Orthogonal Exploration

Dayu Wang[1,2] **Jiaye Yang**[1] **Weikang Li**[1*] **Jiahui Liang**[1] **Yang Li**[1] **Deguo Xia**[1] **Jizhou Huang**[1*]
[1]Baidu Inc. [2]Nanyang Technological University
dayu001@e.ntu.edu.sg yamseyoung@gmail.com
wavejkd@pku.edu.cn {liangjiahui03, liyang164, xiadeguo, huangjizhou01}@baidu.com

Abstract

Large Language Models (LLMs) often suffer from "Reasoning Collapse" on challenging mathematical reasoning tasks, where stochastic sampling produces lexical variations of the same erroneous logic rather than genuine semantic exploration. We observe that failed reasoning traces are often associated with a low-rank bias manifold in the model's hidden-state geometry, which reduces exploration toward corrective solution directions. To address this, we propose **Spectral Orthogonal Exploration (SOE)**, a geometric inference framework under a "Student Guides Teacher" paradigm. Instead of using a weak auxiliary agent for imitation, SOE uses it as an *orthogonal probe* to introduce semantically heterogeneous reasoning signals into the teacher's orthogonal complement of its dominant subspace. This intervention steers the teacher toward more diverse reasoning trajectories and improves exploration beyond standard sampling. Experiments on mathematical benchmarks show that SOE improves average accuracy by **62.4%** and average sampling efficiency by **113.7%** over baseline methods, suggesting that geometric interventions can be effective for mitigating reasoning collapse in mathematical reasoning. We further provide preliminary evidence that SOE is also effective on logic and code generation benchmarks. Code is available at https://github.com/dayuwang401/spectral-orthogonal-exploration.

1 Introduction

Current Large Language Models (LLMs) exhibit a paradox in complex reasoning: while they demonstrate emergent capabilities driven by scaling laws (Kaplan et al., 2020; Wei et al., 2022; Hoffmann et al., 2022; DeepSeek-AI et al., 2025b,a,c; Bubeck et al., 2025), they frequently encounter a recurring failure mode known as *Reasoning Collapse* (Shojaee et al., 2025; Zhao et al., 2025; Guo et al., 2025; Yamin et al., 2025). When facing challenging mathematical reasoning, logic, or code generation tasks, model performance does not decay gracefully but often stagnates abruptly (Petrov et al., 2025). We hypothesize that this phenomenon is associated with a **geometric collapse of the representation space** (Li et al., 2025; Laurent et al., 2023; Fan et al., 2025), as illustrated in Figure 1.

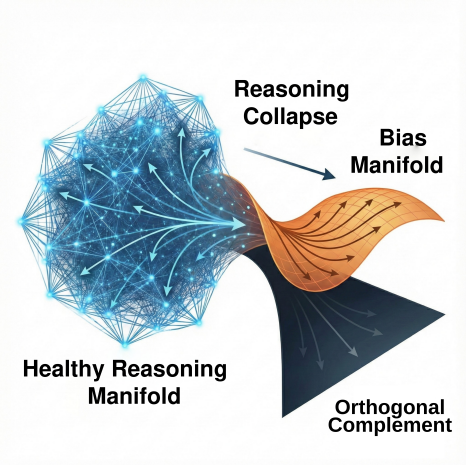
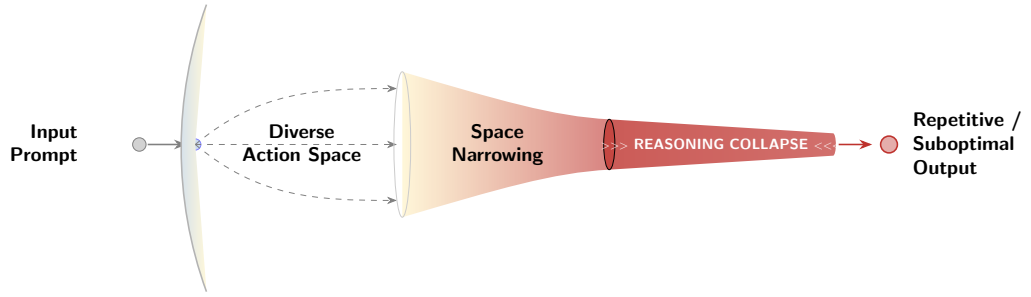


Figure 1: Geometric interpretation of **Reasoning Collapse**. We illustrate reasoning collapse as a transition of the state space from a high-dimensional **Healthy Reasoning Manifold** to a low-rank **Bias Manifold**. This concentration may reduce exploration toward corrective solutions in the **Orthogonal Complement**, as the trajectory becomes confined to a lower-dimensional region.

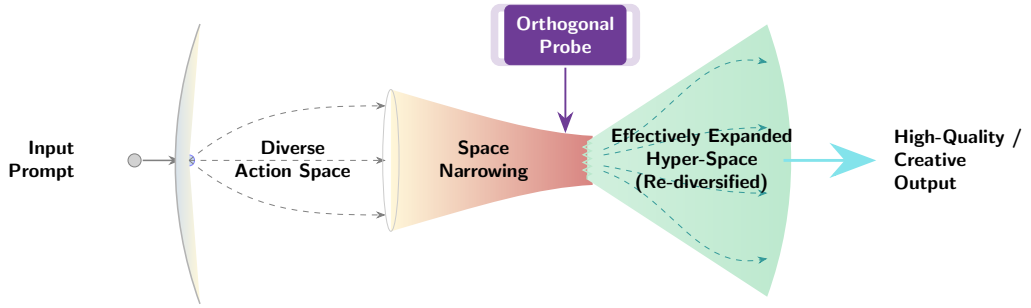
Specifically, we posit the *Low-Rank Manifold Hypothesis*: as a strong model becomes overly confident in an erroneous reasoning path, its internal hidden states may become concentrated in a low-dimensional **Bias Manifold** (Minegishi et al., 2025; Bazarova et al., 2025; Phillips et al., 2025; Park et al., 2025). In this collapsed state, the model's covariance matrix becomes increasingly ill-conditioned, retaining high variance primarily in directions aligned with its current bias, while variance in corrective directions is substantially re-

[*]Project leads and corresponding authors.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 16716–16733



(a) Reasoning Collapse in Vanilla Inference.



(b) Ours: Spectral Orthogonal Exploration (SOE) Intervention.

Figure 2: Mechanism of Spectral Orthogonal Exploration (SOE). (a) Standard training leads to space narrowing and reasoning collapse. (b) To counteract this, we introduce an **Orthogonal Probe** as a geometric intervention. This force effectively disrupts the low-rank confinement and diversifies the reasoning trajectory, expanding the hyper-space to access high-quality solutions.

duced (Dohmatob et al., 2024).

To address this geometric limitation, we introduce the concept of **Spectral Orthogonal Exploration (SOE)**, illustrated in Figure 2b. Our approach is grounded in a fundamental linear algebraic principle: escaping a restricted subspace requires a vector with a non-zero component in its orthogonal complement. We propose a counter-intuitive **Student Guides Teacher** paradigm, where a weaker Student model acts not as a supervisor, but as an **Orthogonal Probe**. Crucially, the Student's utility derives not from its predictive accuracy, but from its structural heterogeneity. Since the Student has not converged toward the same Bias Manifold as the Teacher, its reasoning trajectories—though potentially suboptimal—may retain significant components in directions underexplored by the Teacher. By calculating the **Orthogonal Projection Residual** of the Student's outputs relative to the Teacher's dominant eigenspace, we obtain an intervention signal that steers the Teacher away from its current biased trajectory.

2 Related Work

Traditional alignment and exploration methods, such as high-temperature sampling (Holtzman et al., 2020) or diverse prompting, often struggle to rectify the structural pathology of reasoning collapse (Zhang et al., 2025; Lu et al., 2023; Gudibande et al., 2023; Casper et al., 2023; Ethayarajh, 2019; Shumailov et al., 2023). Because these methods operate primarily in the probabilistic token space, they implicitly assume that lexical diversity equates to semantic exploration (Shi et al., 2025; Shypula et al., 2025).

However, from a geometric perspective, stochastic sampling on a collapsed manifold is similar to **Brownian motion restricted to a 2D plane**: no matter how much noise is injected, the trajectory cannot escape to the third dimension (Seddik et al., 2024; Joseph et al., 2025; Farghly et al., 2025). Consequently, the model generates fluent and diverse yet logically redundant fallacies, remaining effectively trapped within a spectral "blind spot" that stochasticity alone cannot illuminate.

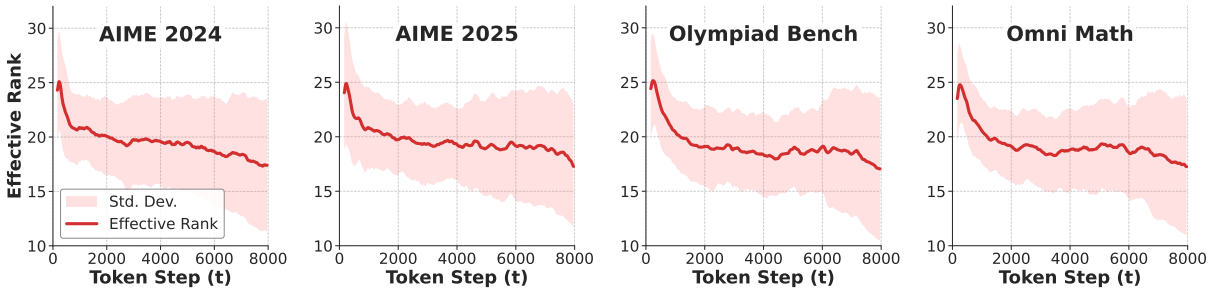


Figure 3: **Universality of Reasoning Collapse across Datasets.** The dynamics of Effective Rank during the reasoning process on AIME 24, AIME 25, Olympiad Bench and Omni-Math. Across all benchmarks, we observe a consistent spectral degeneration pattern: as the token steps increase, the effective rank of the covariance matrix decays, consistent with increasing concentration into a low-dimensional bias manifold across benchmarks.

3 Motivation: The Geometry of Reasoning Collapse

To understand the nature of reasoning failure, we move beyond surface-level token statistics to analyze the geometric properties of the model's representation space. While reasoning collapse is often identified by repetitive text or logical loops, we hypothesize that these patterns may reflect an underlying geometric phenomenon, which we refer to as **Spectral Collapse**.

3.1 The Low-Rank Manifold Hypothesis

Consider the inference process of a Large Language Model as a discrete dynamical system. We define the hidden state trajectory of a reasoning chain up to step t as $H_t = [h_1, h_2, \dots, h_t] \in \mathbb{R}^{t \times d}$, formed by stacking the hidden state vectors. Specifically, the i-th row represents the hidden state $h_i \in \mathbb{R}^{1 \times d}$ at step i. As the model commits to a specific reasoning path, its internal representations may become increasingly concentrated within a lower-dimensional subspace. To describe the dimension of the subspace, we analyze the *Local Covariance Matrix* Σ_t of the recent window of k tokens:

$$\Sigma_t = \frac{1}{k-1} \sum_{i=t-k+1}^{t} (h_i - \mu_t)^\top (h_i - \mu_t) \quad (1)$$

where μ_t is the local mean. When the effective dimensionality of this subspace is low, nearby hidden states become more similar, and the *Local Covariance Matrix* correspondingly exhibits lower effective rank. Based on this, we posit the **Low-Rank Manifold Hypothesis**: during reasoning collapse, the effective rank of Σ_t tends to decrease, indicating that the reasoning process becomes increasingly concentrated in a low-dimensional **Bias Manifold** $\mathcal{M}$ (Gao et al., 2019).

3.2 Quantifying Spectral Collapse

Based on the analysis above, we use the Effective Rank (*EffRank*) (Roy and Vetterli, 2007) to quantify the dimensionality of this bias manifold:

$$\textit{EffRank}(\Sigma_t) = \exp\left(-\sum_{j=1}^{d} \tilde{\sigma}_j \log \tilde{\sigma}_j \right) \quad (2)$$

Here, $\tilde{\sigma}_j = \lambda_j / \sum_{i=1}^{d} \lambda_i$ represents the normalized singular values. We interpret $\textit{EffRank}(\Sigma_t) \ll d$ as evidence of substantial geometric concentration in the representation space. In this state, variance in directions within the orthogonal complement of the dominant subspace is substantially reduced. Consequently, even with stochastic sampling (temperature $T > 0$), generated trajectories may remain concentrated within the Bias Manifold, reducing exploration toward corrective solutions in the orthogonal complement.

3.3 Empirical Analysis

To examine this hypothesis, we conducted a pilot study across four challenging mathematical benchmarks: AIME 24 (Patel et al., 2024), AIME 25 (Petrov et al., 2025), Olympiad Bench (He et al., 2024) and Omni-Math (Gao et al., 2024). We use `Qwen3-4B-Instruct-2507` (Team, 2025b) to sample reasoning chains with a high temperature ($T = 0.7$). To specifically isolate the phenomenon of reasoning collapse, we curate a subset of trajectories that are both factually incorrect and excessively verbose (exceeding 4,000 tokens). For each selected trajectory, we computed the Effective Rank of the covariance matrix of the final hidden states using a sliding window of size $k = 64$.

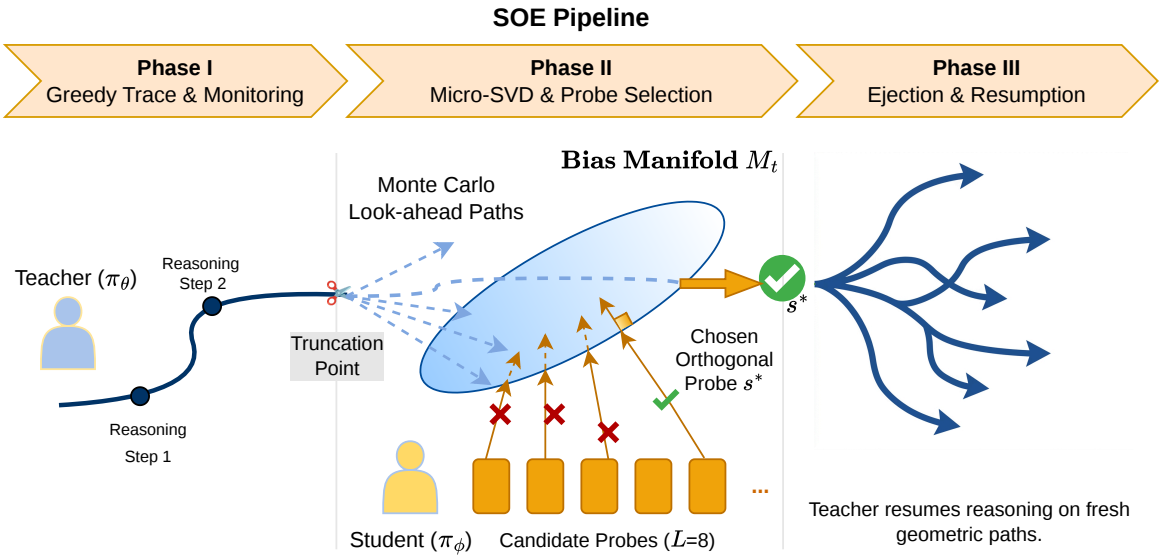


Figure 4: The pipeline of Spectral Orthogonal Exploration (SOE). A greedy reasoning trace from the Teacher is first monitored and truncated at candidate collapse points. At each truncation point, Monte Carlo samples are used to estimate the Teacher's local bias manifold via Micro-SVD, and the Student generates multiple short candidate probes. The probe with the largest orthogonal residual in the Teacher's latent space is selected and stitched into the context. The Teacher then resumes reasoning from the modified prefix, thereby exploring geometric directions outside its original bias manifold.

As illustrated in Figure 3, we observe a substantial decrease in effective rank over the course of long, erroneous reasoning chains. Despite a theoretical full rank of $d = 2560$, the effective dimensionality consistently decays as generation progresses. This quantitative evidence is consistent with our hypothesis that reasoning loops are associated with concentration into a low-dimensional bias manifold, which may reduce the geometric degrees of freedom available for divergent exploration. These observations motivate the use of a geometric intervention to encourage broader exploration.

To further substantiate these findings, we provide an extended analysis in Appendix B. There, we provide additional evidence that similar collapse patterns can be observed across multiple model sizes.

4 Methodology: The SOE Framework

Motivated by the observed geometric decay, we develop the **Spectral Orthogonal Exploration (SOE)** framework. This section describes the SOE algorithm for identifying reasoning collapse and intervening during the inference process.

4.1 Overview of the Geometric Interaction

The framework operates in a post-hoc intervention setting. As shown in Figure 4, the process consists of three stages:

1. **Greedy Trace and Monitoring:** We first employ the Teacher π_θ to generate a complete reasoning chain via greedy decoding. We then collect trajectories that yield incorrect answers, treating these erroneous paths as candidate instances of collapsed reasoning. These trajectories are partitioned at critical reasoning milestones (e.g., the onset of logical steps). This procedure yields a set of prefix trajectories, allowing us to probe the model's latent state at various depths of the reasoning process.

2. **Micro-SVD and Probe Selection:** At each truncation point t_i, we estimate the local bias manifold $\mathcal{M}_t$ using Monte Carlo look-ahead simulations and Micro-SVD to identify the dominant subspace associated with the Teacher's current reasoning trajectory. For each truncation point, the Student π_ϕ generates candidate probe sequences (fixed at $L = 8$ tokens). We calculate the orthogonality of these candidates relative to the specific manifold $\mathcal{M}_t$ at that step, selecting the optimal probe s_t^* that maximizes the orthogonality score.

3. **Intervention and Resumption:** The previ-

ous process yields distinct intervened contexts (the truncated histories stitched with their respective orthogonal probes). The Teacher π_θ then resumes reasoning via sampling from each of these modified contexts, allowing it to continue exploration from three different geometric intervention points.

A detailed example of the algorithm is provided in Appendix A.

4.2 Manifold Estimation

To estimate the bias manifold of the Teacher, we assume the Teacher's internal representations at truncation point t follow a local distribution with covariance $\mathbf{\Sigma_t}$. Our goal is to obtain an *unbiased estimate* of the principal components of $\mathbf{\Sigma_t}$.

Unbiased Covariance Estimator. Given the context $x_{<t}$ at the truncation point t, we employ a Monte Carlo method to sample N stochastic look-ahead trajectories $\{\tau_1, \ldots, \tau_N\}$ from the Teacher $\pi_\theta(\cdot|x_{<t})$. Let $\mathbf{h}_i \in \mathbb{R}^d$ be the aggregated hidden state of the i-th trajectory. We calculate the sample mean $\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}_i$ and construct the centered state matrix $\mathbf{H} \in \mathbb{R}^{d\times N}$:

$$\mathbf{H} = [\mathbf{h}_1 - \hat{\boldsymbol{\mu}}, \ldots, \mathbf{h}_N - \hat{\boldsymbol{\mu}}] \tag{3}$$

Theoretically, the sample covariance matrix $\hat{\mathbf{C}} = \frac{1}{N-1}\mathbf{H}\mathbf{H}^T$ serves as an unbiased estimator of the true population covariance $\mathbf{\Sigma_t}$. By analyzing the spectrum of $\hat{\mathbf{C}}$, we estimate the directions of maximum variance associated with the Teacher's current reasoning trajectory.

The Micro-SVD Algorithm. Directly computing eigenvectors for $\hat{\mathbf{C}} \in \mathbb{R}^{d\times d}$ is computationally prohibitive. We leverage the SVD duality by computing the Gram matrix $\mathbf{G} \in \mathbb{R}^{N\times N}$ of our Monte Carlo samples:

$$\mathbf{G} = \mathbf{H}^T\mathbf{H} \tag{4}$$

Solving the eigen-problem for the smaller matrix $\mathbf{G}\mathbf{v}_k = \lambda_k\mathbf{v}_k$ allows us to recover the principal components $\mathbf{u}_k$ of the Teacher's local subspace:

$$\mathbf{u}_k = \frac{1}{\sqrt{\lambda_k}}\mathbf{H}\mathbf{v}_k \tag{5}$$

We define the **Bias Manifold** $\mathcal{M}_t$ as the subspace spanned by the top-k eigenvectors $\mathbf{U}_\parallel = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. Directions in the orthogonal complement of this subspace provide candidate directions for intervention and exploration.

4.3 Probe Selection: Orthogonal Latent Stitching (OLS)

With the manifold $\mathcal{M}_t$ identified, we employ Orthogonal Latent Stitching to intervene on the model's current reasoning trajectory.

Heterogeneous Probing. The Student model π_ϕ generates a set of candidate sequences $\mathcal{C}_{Student} = \{s_1, \ldots, s_M\}$, each with a fixed length of $L = 8$ tokens. These short sequences act as geometric probe vectors. To evaluate their geometry, we map them into the Teacher's latent space via a forward pass: $\mathbf{z}_j = \mathit{Forward}_{\pi_\theta}(s_j \mid x_{<t})$.

Residual Maximization. We compute the projection of each candidate $\mathbf{z}_j$ onto the Teacher's bias manifold using the projection operator $\mathbf{P}_\parallel = \mathbf{U}_\parallel\mathbf{U}_\parallel^T$. The orthogonal residual $\mathbf{r}_j$ represents the component of the Student's probe that lies in the orthogonal complement of the Teacher's dominant subspace:

$$\mathbf{r}_j = (\mathbf{I} - \mathbf{P}_\parallel)(\mathbf{z}_j - \hat{\boldsymbol{\mu}}) \tag{6}$$

We select the candidate s^* that maximizes the Orthogonality Score, defined as the normalized energy of this residual, where $\epsilon > 0$ is a small constant introduced to avoid division by zero:

$$s^* = \arg\max_{s_j \in \mathcal{C}_{Student}} \frac{\|\mathbf{r}_j\|_2}{\|\mathbf{z}_j - \hat{\boldsymbol{\mu}}\|_2 + \epsilon} \tag{7}$$

This orthogonal token sequence s^* is then stitched to the truncated context, providing an intervention that steers the Teacher toward exploration from a new geometric direction outside its original bias manifold.

5 Experiments

To validate the theoretical claims of Spectral Orthogonal Exploration (SOE), we conducted extensive experiments focused on high-difficulty mathematical reasoning tasks.

5.1 Experimental Setup

Models. We adopt a "Weak-Student and Strong-Teacher" configuration to validate our proposal: We use `Qwen3-4B-Instruct-2507` as the Teacher, and use `Gemma-3-4B-IT` (Team, 2025a) as the Student. Crucially, the Student's different architecture and training data ensure that its bias manifold is not perfectly aligned with the Teacher's.

Method	AIME 24	AIME 25	MATH-500	Olympiad Bench	Omni-Math (Hard)	Average
Baseline (Self-Consistency)	38.5%	35.3%	33.7%	11.7%	14.5%	26.7%
Ours (SOE)	**76.9%**	**70.6%**	**45.9%**	**15.5%**	**20.8%**	**45.9%**
Relative Improvement	+99.7%	+100.0%	+36.2%	+32.5%	+43.4%	+62.4%

Table 1: Pass@16 accuracy on the **Difficult Subset** (problems where Teacher's greedy decoding failed). SOE demonstrates a consistent improvement over the Baseline, confirming that directed orthogonal exploration is superior to standard self-consistency sampling. All the results are sampled under identical conditions, with the same prompt, $temperature = 0.7$ and a maximum context (sampling) length capped at 8192 tokens.

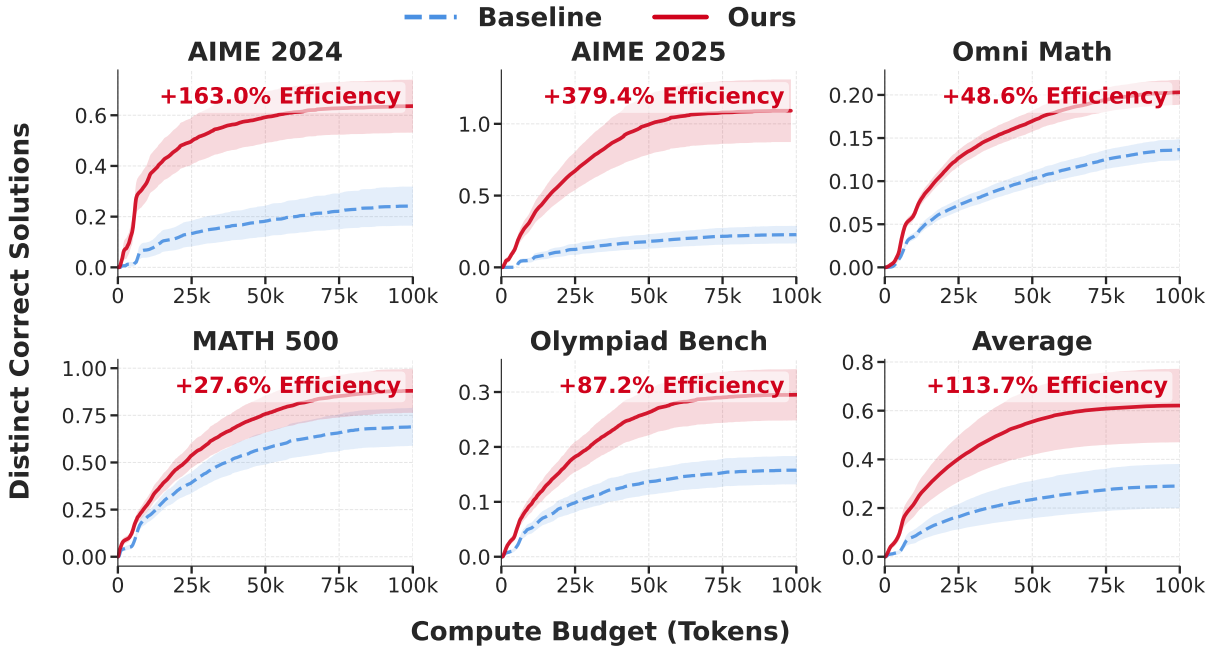


Figure 5: Semantic Exploration Efficiency. The curves depict the cumulative number of distinct correct solutions discovered as a function of the compute budget (total tokens). SOE significantly outperforms the Baseline by maintaining a near-linear discovery rate, whereas the Baseline exhibits logarithmic saturation.

Dataset	Teacher	Student
AIME 2024	**50.0%**	13.3%
AIME 2025	**33.7%**	6.7%
MATH-500	**86.8%**	71.2%
Olympiad Bench	**45.3%**	27.7%
Omni-Math(Hard)	**18.3%**	4.2%

Table 2: Comparison of fundamental capabilities between Teacher and Student models on mathematical benchmarks. Both the Teacher and Student are evaluated under identical conditions, with the same prompt, greedy decoding and a maximum context length capped at 8192 tokens.

Datasets. We evaluate performance on five challenging mathematical benchmarks: AIME 2024, AIME 2025, MATH-500 (Hendrycks et al., 2021), Olympiad Bench and Omni-Math. Given the strong capabilities of the Teacher model, we specifically selected problems from Omni-Math with a difficulty level greater than or equal to 8.

Evaluation. For all benchmarks, model outputs are post-processed using a combination of regular-expression-based normalization and the `MathEvaluator` library to extract and verify final answers against ground truth.

Baselines and Implementation Details. We compare two experimental settings:

- **Baseline (Self-Consistency):** Standard sampling with the Teacher model ($T = 0.7$). This represents the control group for stochastic exploration within the Bias Manifold.
- **Ours (SOE):** We generate $N = 8$ candidates from the Student ($T = 1.0$), compute their

Teacher	Student	Benchmark	Baseline	Ours (SOE)
Qwen3-4B-Instruct-2507	DeepSeek-R1-Distill-Qwen-7B	AIME 2024	38.46%	**76.92%**
Qwen3-4B-Instruct-2507	DeepSeek-R1-Distill-Qwen-7B	AIME 2025	35.29%	**64.70%**
Qwen3-4B-Instruct-2507	Mistral-7B-Instruct-v0.3	AIME 2024	38.46%	**69.23%**
Qwen3-4B-Instruct-2507	Mistral-7B-Instruct-v0.3	AIME 2025	35.29%	**58.82%**
Qwen3-8B	Gemma-3-4B-IT	AIME 2024	14.29%	**42.86%**
Qwen3-8B	Gemma-3-4B-IT	AIME 2025	4.35%	**26.09%**
Qwen3-32B	Gemma-3-4B-IT	AIME 2024	11.11%	**55.56%**
Qwen3-32B	Gemma-3-4B-IT	AIME 2025	21.74%	**34.78%**

Table 3: Cross-family and scale generalization of SOE. The method remains effective with heterogeneous Teacher–Student pairs and larger Teacher models.

embeddings using the Teacher model, and mechanically stitch the candidate maximizing the Orthogonality Score ($\Omega(s) \to 1$) into the Teacher's context. The subsequent reasoning chain is completed by the Teacher ($T = 0.7$).

5.2 Main Results

We evaluate the solution accuracy using **Pass@16**. As shown in Table 1, SOE significantly outperforms the Self-Consistency baseline across all datasets on the difficult subset.

We further compare SOE against a strong step-level Best-of-N baseline using a Process Reward Model (PRM). As shown in Table 4, SOE consistently surpasses the PRM baseline on AIME 2024, AIME 2025, and MATH-500 under matched sampling settings, indicating that the gains are not solely attributable to stronger step selection.

Dataset	PRM Best-of-N	Ours (SOE)	Rel. Imp.
AIME 2024	69.23%	**76.90%**	+11.08%
AIME 2025	58.82%	**70.60%**	+20.03%
MATH-500	40.98%	**45.90%**	+12.01%

Table 4: Comparison against a strong step-level Best-of-N baseline using a Process Reward Model (PRM). All sampling positions and the number of subsequent trajectories are matched to SOE.

To evaluate generalizability beyond a single 4B Teacher–Student pair, we further test heterogeneous cross-family combinations and larger Teacher models. Table 3 shows that SOE remains effective when using DeepSeek-R1-Distill-Qwen-7B and Mistral-7B-Instruct-v0.3 (Jiang et al., 2023) as Student models, and when scaling the Teacher from Qwen3-4B to Qwen3-8B and Qwen3-32B.

5.3 Control Experiments on Rank Decay

To disentangle rank decay from mere verbosity or looping, we conduct matched-control experiments on AIME 2025 with 16 samples per problem. Specifically, we compare the percentage drop in Effective Rank between the first 500 tokens and the final 500 tokens across four cohorts: short/long and correct/wrong traces.

Trace Category	Avg. Rank Drop
Short & Correct ($<$ 4k tokens)	4.82%
Short & Wrong ($<$ 4k tokens)	19.65%
Long & Correct ($>$ 5k tokens)	13.14%
Long & Wrong ($>$ 5k tokens)	27.04%

Table 5: Matched-control analysis on AIME 2025. Effective-rank drop is computed by comparing the first 500 tokens and the final 500 tokens of each trajectory.

As shown in Table 5, long incorrect traces suffer more than double the rank decay of long correct traces, while short incorrect traces already exhibit a substantial drop even without extended repetition. These results suggest that rank decay is not merely a byproduct of verbosity, but is closely associated with reasoning failure.

5.4 Exploration Efficiency

Given that Spectral Orthogonal Exploration (SOE) introduces additional inference cost, it is important to assess whether this overhead translates into more effective exploration. To this end, we compare SOE with the Self-Consistency baseline in terms of **Semantic Exploration Efficiency**, which measures how efficiently a method discovers *semantically distinct* correct reasoning trajectories under a fixed generation budget. This metric is particularly relevant for **data synthesis**, where semantically diverse correct traces are more useful than redundant paraphrases of the same solution.

Metric Definition. We use `jina` embeddings (Günther et al., 2023) to represent generated correct solution traces. A correct solution is counted as "new" only if its cosine similarity to all previously discovered correct solutions is below a threshold of

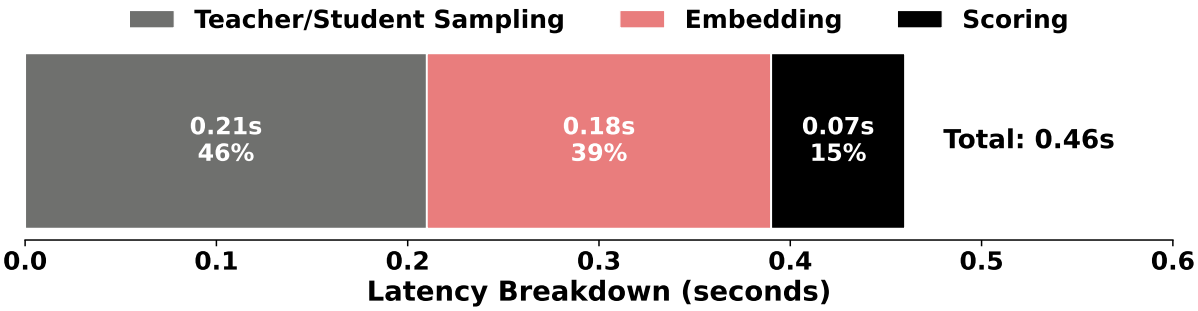


Figure 6: Latency breakdown of a single SOE intervention step evaluated on the AIME 2025 benchmark. The process was measured on an NVIDIA A800 GPU using the vLLM pipeline.

Benchmark	Baseline	Ours	Ours (Time-Norm.)
AIME 2024	42.86%	77.14%	68.39%
AIME 2025	36.00%	72.00%	63.83%
MATH-500	35.12%	44.64%	39.57%
Olympiad Bench	9.49%	14.09%	12.49%
Omni-Math	14.39%	20.45%	18.13%

Table 6: Sample accuracy under wall-clock normalization. Following the measured 12.8% runtime overhead of SOE under vLLM, we divide the SOE count by 1.128 for a strict time-normalized comparison.

$\tau = 0.95$, thereby filtering out trivial paraphrases.

As illustrated in Figure 5, SOE consistently achieves higher exploration efficiency than the baseline.

As detailed in Figure 6, under an optimized `vLLM` deployment, a single SOE intervention incurs only about 0.46s overhead, including approximately 0.21s for Teacher/Student sampling, 0.18s for embedding extraction, and 0.07s for scoring. Given an average end-to-end inference time of roughly 3.6s per problem in our setting, this corresponds to only a 12.8% runtime overhead. Nevertheless, Table 6 shows that SOE still achieves a higher density of correct solutions per unit wall-clock time than standard Self-Consistency across all five benchmarks.

Additional analyses, including orthogonality-score distributions, threshold-sensitivity results, and comprehensive ablations across components and model scales, are deferred to Appendix C.2 and Appendix C.

5.5 Preliminary Results on Logic and Code Generation

To further probe whether SOE extends beyond mathematical reasoning, we additionally evaluate it on ZebraLogic (Lin et al., 2025) and HumanEvalPlus (Liu et al., 2023). As shown in Table 7, SOE also improves problem-level accuracy on both tasks, suggesting that the benefit of orthogonal exploration is not limited to math benchmarks.

Benchmark	Baseline	Ours (SOE)	Rel. Imp.
ZebraLogic	56.23%	**58.72%**	+4.43%
HumanEvalPlus	10.00%	**16.67%**	+66.7%

Table 7: Preliminary evaluation of SOE beyond mathematical reasoning.

6 Analysis: Geometric Perspectives on Reasoning Collapse

Based on our analysis, we view "Reasoning Collapse" as being associated with *rank concentration* in the hidden-state covariance matrix.

In this section, we discuss two questions: (1) why LLM reasoning trajectories may become increasingly concentrated in low-dimensional subspaces, and (2) why our Orthogonal Latent Stitching (OLS) can act as a geometric intervention rather than simply adding new factual content.

6.1 The Trap of Low-Rank Self-Consistency

Let $H_t \in \mathbb{R}^{N \times d}$ be the matrix of hidden states in the context window at step t. We define the local covariance matrix $\Sigma_t = \frac{1}{N-1} H_t^T H_t$. Reasoning collapse is associated with a decrease in the effective rank of Σ_t, i.e., $\mathrm{EffRank}(\Sigma_t) \ll d$, indicating that the states become increasingly concentrated in a subspace $\mathcal{S}$.

Proposition 1 (Invariance of Subspace). *If the input context lies within a low-rank subspace $\mathcal{S}$, the output of a standard Self-Attention layer lies within the same subspace (or a linear transform thereof), limiting spontaneous exploration of its orthogonal complement.* (Dong et al., 2021; Wang et al., 2020)

The detailed proof and an analysis of why non-linearities (FFNs) may fail to substantially restore dimensionality are provided in Appendix D.1.

6.2 Theoretical Justification for Progressive Rank Decay

Why does the effective rank decrease over time during long chain-of-thought reasoning? We relate this phenomenon to a **spectrum contraction** effect of the Softmax operator in high-confidence regimes (Yang et al., 2017).

Proposition 2 (Spectrum Contraction). *As the model becomes more confident (the norm of logits increases), the attention mechanism satisfies a non-expansion-style constraint on the dimensionality of the resulting representation, which can contribute to spectral concentration in the covariance matrix.*

Please refer to Appendix D.2 for the formal proof. This result provides a theoretical perspective on why reasoning trajectories may become progressively concentrated over time, reducing exploration in directions useful for error correction.

6.3 Orthogonal Injection: Geometry and Attention Rerouting

A critical question arises: does SOE work because the Student provides *new knowledge*, or because it introduces a geometric perturbation that changes the Teacher's subsequent trajectory? Our view is that the evidence more strongly supports the latter interpretation.

Consider the contribution to the spectrum in a direction $u_\perp$ lying in the orthogonal complement of the dominant subspace. Prior to injection, the variance in this direction is $\sigma_\perp^2 \approx 0$. After injecting an orthogonal vector $v_\perp$ (where $\|v_\perp\| > 0$), the covariance matrix acquires non-zero energy in this direction, yielding

$$\mathrm{rank}(\Sigma_{\text{new}}) \geq \mathrm{rank}(\Sigma_{\text{old}}) \tag{8}$$

and, in the idealized full-rank-augmentation case,

$$\mathrm{rank}(\Sigma_{\text{new}}) = \mathrm{rank}(\Sigma_{\text{old}}) + 1. \tag{9}$$

This change does not by itself guarantee better reasoning, but it provides a geometric mechanism by which the trajectory can access previously underexplored directions. In practice, we hypothesize that semantically structured Student probes are helpful not merely because they add tokens, but because they introduce heterogeneous signals that reroute the Teacher's attention away from its current biased trajectory. From this perspective, SOE can be understood as a geometric intervention that expands the space of subsequent exploration, rather than purely as a knowledge-injection mechanism.

7 Conclusion

We presented evidence that *Reasoning Collapse* is associated with geometric concentration into a low-rank *Bias Manifold*, which can reduce exploration toward corrective directions in the orthogonal complement. To address this, we introduced **Spectral Orthogonal Exploration (SOE)**, a "Student Guides Teacher" framework that uses *Orthogonal Latent Stitching* to intervene on biased reasoning trajectories. Our empirical and theoretical analyses suggest that SOE is effective partly because it expands the geometric space available for subsequent exploration, rather than solely because it injects additional information. More broadly, these findings suggest that mitigating reasoning bottlenecks may benefit from going beyond probabilistic sampling and incorporating targeted geometric interventions in latent space.

8 Limitations

While our proposed Spectral Orthogonal Exploration (SOE) framework demonstrates significant efficacy in mitigating reasoning collapse, we acknowledge several limitations that merit further investigation.

First, although SOE improves sample efficiency by requiring fewer total tokens to discover correct solutions, it still introduces additional wall-clock latency due to the Monte Carlo look-ahead, hidden-state extraction, and scoring steps required to estimate the local bias manifold and select orthogonal probes.

Second, the core mechanism of SOE relies on explicitly accessing and manipulating the model's internal hidden states to compute effective rank and orthogonal residuals. This requirement restricts the applicability of our method to open-weights models, precluding its direct use with closed-source LLMs where only API access is available.

Finally, although we include additional validation on larger Teacher models and preliminary experiments beyond mathematics, our empirical study remains centered on mathematical reasoning, and a full-scale evaluation across broader domains and model families is still needed.

References

Alexandra Bazarova, Aleksandr Yugay, Andrey Shulga, Alina Ermilova, Andrei Volodichev, Konstantin Polev, Julia Belikova, Rauf Parchiev, Dmitry Simakov, Maxim Savchenko, Andrey Savchenko, Serguei Barannikov, and Alexey Zaytsev. 2025. Hallucination detection in llms with topological divergence on attention graphs. *Preprint*, arXiv:2504.10063.

Sébastien Bubeck, Christian Coester, Ronen Eldan, Timothy Gowers, Yin Tat Lee, Alexandru Lupsasca, Mehtaab Sawhney, Robert Scherrer, Mark Sellke, Brian K. Spears, Derya Unutmaz, Kevin Weil, Steven Yin, and Nikita Zhivotovskiy. 2025. Early science acceleration experiments with gpt-5. *Preprint*, arXiv:2511.16072.

Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémy Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, and 1 others. 2023. Open problems and fundamental limitations of reinforcement learning from human feedback. *arXiv preprint arXiv:2307.15217*.

DeepSeek-AI, Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, and 181 others. 2025a. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

DeepSeek-AI, Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, and 181 others. 2025b. Deepseek-v3 technical report. *Preprint*, arXiv:2412.19437.

DeepSeek-AI, Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenhao Xu, Chong Ruan, Damai Dai, Daya Guo, Dejian Yang, and 245 others. 2025c. Deepseek-v3.2: Pushing the frontier of open large language models. *Preprint*, arXiv:2512.02556.

Elvis Dohmatob, Yunzhen Feng, and Julia Kempe. 2024. Model collapse demystified: The case of regression. *Preprint*, arXiv:2402.07712.

Yihe Dong, Jean-Baptiste Cordonnier, and Andreas Loukas. 2021. Attention is not all you need: Pure attention loses rank doubly exponentially with depth. In *International conference on machine learning*, pages 2793–2803. PMLR.

Kawin Ethayarajh. 2019. How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings. *arXiv preprint arXiv:1909.00512*.

Ziqing Fan, Siyuan Du, Shengchao Hu, Pingjie Wang, Li Shen, Ya Zhang, Dacheng Tao, and Yanfeng Wang. 2025. Combatting dimensional collapse in llm pretraining data via diversified file selection. *Preprint*, arXiv:2504.20644.

Tyler Farghly, Peter Potaptchik, Samuel Howard, George Deligiannidis, and Jakiw Pidstrigach. 2025. Diffusion models and the manifold hypothesis: Log-domain smoothing is geometry adaptive. *Preprint*, arXiv:2510.02305.

Bofei Gao, Feifan Song, Zhe Yang, Zefan Cai, Yibo Miao, Qingxiu Dong, Lei Li, Chenghao Ma, Liang Chen, Runxin Xu, Zhengyang Tang, Benyou Wang, Daoguang Zan, Shanghaoran Quan, Ge Zhang, Lei Sha, Yichang Zhang, Xuancheng Ren, Tianyu Liu, and Baobao Chang. 2024. Omni-math: A universal olympiad level mathematic benchmark for large language models. *Preprint*, arXiv:2410.07985.

Jun Gao, Di He, Xu Tan, Tao Qin, Liwei Wang, and Tie-Yan Liu. 2019. Representation degeneration problem in training natural language generation models. *arXiv preprint arXiv:1907.12009*.

Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. 2021. Transformer feed-forward layers are key-value memories. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5484–5495.

Arnav Gudibande, Eric Wallace, Charlie Snell, Xinyang Geng, Hao Liu, Pieter Abbeel, Sergey Levine, and Dawn Song. 2023. The false promise of imitating proprietary llms. *arXiv preprint arXiv:2305.15717*.

Dadi Guo, Jiayu Liu, Zhiyuan Fan, Zhitao He, Haoran Li, Yuxin Li, Yumeng Wang, and Yi R. Fung. 2025. Mathematical proof as a litmus test: Revealing failure modes of advanced large reasoning models. *Preprint*, arXiv:2506.17114.

Michael Günther, Jackmin Ong, Isabelle Mohr, Alaeddine Abdessalem, Tanguy Abel, Mohammad Kalim Akram, Susana Guzman, Georgios Mastrapas, Saba Sturua, Bo Wang, Maximilian Werk, Nan Wang, and Han Xiao. 2023. Jina embeddings 2: 8192-token general-purpose text embeddings for long documents. *Preprint*, arXiv:2310.19923.

Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, and 1 others. 2024. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3828–3850.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, and 3 others. 2022. Training compute-optimal large language models. *Preprint*, arXiv:2203.15556.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2020. The curious case of neural text degeneration. *Preprint*, arXiv:1904.09751.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b. *Preprint*, arXiv:2310.06825.

Federico Arangath Joseph, Jerome Sieber, Melanie N. Zeilinger, and Carmen Amo Alonso. 2025. Lambda-skip connections: the architectural component that prevents rank collapse. *Preprint*, arXiv:2410.10609.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Thomas Laurent, James H. von Brecht, and Xavier Bresson. 2023. Feature collapse. *Preprint*, arXiv:2305.16162.

Melody Zixuan Li, Kumar Krishna Agrawal, Arna Ghosh, Komal Kumar Teru, Adam Santoro, Guillaume Lajoie, and Blake A. Richards. 2025. Tracing the representation geometry of language models from pretraining to post-training. *Preprint*, arXiv:2509.23024.

Bill Yuchen Lin, Ronan Le Bras, Kyle Richardson, Ashish Sabharwal, Radha Poovendran, Peter Clark, and Yejin Choi. 2025. Zebralogic: On the scaling limits of llms for logical reasoning. *arXiv preprint arXiv:2502.01100*.

Jiawei Liu, Chunqiu Steven Xia, Yuyao Wang, and Lingming Zhang. 2023. Is your code generated by chatgpt really correct? rigorous evaluation of large language models for code generation. *Advances in neural information processing systems*, 36:21558–21572.

Pan Lu, Liang Qiu, Wenhao Yu, Sean Welleck, and Kai-Wei Chang. 2023. A survey of deep learning for mathematical reasoning. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers)*, pages 14605–14631.

Gouki Minegishi, Hiroki Furuta, Takeshi Kojima, Yusuke Iwasawa, and Yutaka Matsuo. 2025. Topology of reasoning: Understanding large reasoning models through reasoning graph properties. *Preprint*, arXiv:2506.05744.

Seongheon Park, Xuefeng Du, Min-Hsuan Yeh, Haobo Wang, and Yixuan Li. 2025. Steer llm latents for hallucination detection. *Preprint*, arXiv:2503.01917.

Bhrij Patel, Souradip Chakraborty, Wesley A. Suttle, Mengdi Wang, Amrit Singh Bedi, and Dinesh Manocha. 2024. Aime: Ai system optimization via multiple llm evaluators. *Preprint*, arXiv:2410.03131.

Ivo Petrov, Jasper Dekoninck, Lyuben Baltadzhiev, Maria Drencheva, Kristian Minchev, Mislav Balunović, Nikola Jovanović, and Martin Vechev. 2025. Proof or bluff? evaluating llms on 2025 usa math olympiad. *Preprint*, arXiv:2503.21934.

Edward Phillips, Sean Wu, Soheila Molaei, Danielle Belgrave, Anshul Thakur, and David Clifton. 2025. Geometric uncertainty for detecting and correcting hallucinations in llms. *Preprint*, arXiv:2509.13813.

Olivier Roy and Martin Vetterli. 2007. The effective rank: A measure of effective dimensionality. In *2007 15th European signal processing conference*, pages 606–610. IEEE.

Mohamed El Amine Seddik, Suei-Wen Chen, Soufiane Hayou, Pierre Youssef, and Merouane Debbah. 2024. How bad is training on synthetic data? a statistical analysis of language model collapse. *Preprint*, arXiv:2404.05090.

Weijie Shi, Yue Cui, Yaguang Wu, Jingzhi Fang, Shibo Zhang, Mengze Li, Sirui Han, Jia Zhu, Jiajie Xu, and Xiaofang Zhou. 2025. Semantic-guided diverse decoding for large language model. *Preprint*, arXiv:2506.23601.

Parshin Shojaee, Iman Mirzadeh, Keivan Alizadeh, Maxwell Horton, Samy Bengio, and Mehrdad Farajtabar. 2025. The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. *Preprint*, arXiv:2506.06941.

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson. 2023. The curse of recursion: Training on generated data makes models forget. *arXiv preprint arXiv:2305.17493*.

Alexander Shypula, Shuo Li, Botong Zhang, Vishakh Padmakumar, Kayo Yin, and Osbert Bastani. 2025. Evaluating the diversity and quality of llm generated content. *Preprint*, arXiv:2504.12522.

Gemma Team. 2025a. Gemma 3.

Qwen Team. 2025b. Qwen3 technical report. *Preprint*, arXiv:2505.09388.

Qwen Team. 2025c. Qwq-32b: Embracing the power of reinforcement learning.

Sinong Wang, Belinda Z Li, Madian Khabsa, Han Fang, and Hao Ma. 2020. Linformer: Self-attention with linear complexity. *arXiv preprint arXiv:2006.04768*.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. 2022. Emergent abilities of large language models. *Preprint*, arXiv:2206.07682.

Khurram Yamin, Shantanu Gupta, Gaurav R. Ghosal, Zachary C. Lipton, and Bryan Wilder. 2025. Failure modes of llms for causal reasoning on narratives. *Preprint*, arXiv:2410.23884.

Zhilin Yang, Zihang Dai, Ruslan Salakhutdinov, and William W Cohen. 2017. Breaking the softmax bottleneck: A high-rank rnn language model. *arXiv preprint arXiv:1711.03953*.

Yue Zhang, Yafu Li, Leyang Cui, Deng Cai, Lemao Liu, Tingchen Fu, Xinting Huang, Enbo Zhao, Yu Zhang, Yulong Chen, and 1 others. 2025. Siren's song in the ai ocean: A survey on hallucination in large language models. *Computational Linguistics*, pages 1–46.

Chengshuai Zhao, Zhen Tan, Pingchuan Ma, Dawei Li, Bohan Jiang, Yancheng Wang, Yingzhen Yang, and Huan Liu. 2025. Is chain-of-thought reasoning of llms a mirage? a data distribution lens. *Preprint*, arXiv:2508.01191.

A Detailed Example of Spectral Orthogonal Exploration

In this section, we provide a step-by-step walk-through of the SOE pipeline on a specific problem from the AIME 2025 dataset. We visualize the breakdown of the Teacher's reasoning collapse and how the Student's orthogonal probe successfully ejects the generation from the Bias Manifold.

A.1 Problem Input

Input Prompt

System Prompt: Please reason step by step, and put your final answer within `\boxed{}`.
User Problem: In $triangleABC$ points D and E lie on $overlineAB$ so that $AD < AE < AB$, while points F and G lie on $overlineAC$ so that $AF < AG < AC$. Suppose $AD = 4$, $DE = 16$, $EB = 8$, $AF = 13$, $FG = 52$, and $GC = 26$. Let M be the reflection of D through F, and let N be the reflection of G through E. The area of quadrilateral $DEGF$ is 288. Find the area of heptagon $AFNBCEM$.

A.2 Baseline: Reasoning Collapse (Greedy Decoding)

The Teacher model (π_θ) initially fails to solve the problem. Following is the generation.

We are given:

- Points D, E on AB such that $AD < AE < AB$,
- Points F, G on AC such that $AF < AG < AC$,
- Lengths:
  - $AD = 4$
  - $DE = 16$
  - $EB = 8$
  - $AF = 13$
  - $FG = 52$
  - $GC = 26$
- M is the reflection of D through F,
- N is the reflection of G through E,
- The area of quadrilateral $DEGF$ is 288,
- Find the area of the heptagon $AFNBCEM$.

Step 1: Understand the given structure

On segment AB:

$$AD = 4, \quad DE = 16, \quad EB = 8$$

$$AB = AD + DE + EB = 4 + 16 + 8 = 30$$

On segment AC:

$$AF = 13, \quad FG = 52, \quad GC = 26$$

$$AC = AF + FG + GC = 13 + 52 + 26 = 91$$

Step 2: Understanding points M and N

- M is the reflection of D through point F, meaning F is the midpoint of segment DM.
- N is the reflection of G through point E, meaning E is the midpoint of segment GN.

Since D lies on AB while F lies on AC, the reflection is a point reflection in the plane, not along a line. Similarly, N need not lie on any side of triangle ABC.

The vertices of the required heptagon are

$$A, F, N, B, C, E, M.$$

Step 3: Strategy

We aim to express the area of the heptagon $AFNBCEM$ in terms of the area of triangle ABC. However, since points M and N lie outside triangle ABC, the heptagon extends beyond the triangle, and we cannot simply subtract interior regions.

Thus, we adopt a vector-area approach.

Step 4: Vector setup

Let

$$\vec{AB} = \vec{b}, \qquad \vec{AC} = \vec{c}.$$

Then:

$$|\vec{b}\times\vec{c}| = AB\cdot AC\cdot\sin\theta = 30\cdot 91\cdot\sin\theta = 2730\sin\theta.$$

Points along AB:

$$\vec{D} = \frac{4}{30}\vec{b} = \frac{2}{15}\vec{b}, \qquad \vec{E} = \frac{20}{30}\vec{b} = \frac{2}{3}\vec{b}.$$

Points along AC:

$$\vec{F} = \frac{13}{91}\vec{c} = \frac{1}{7}\vec{c}, \qquad \vec{G} = \frac{65}{91}\vec{c} = \frac{5}{7}\vec{c}.$$

Step 5: Area of quadrilateral $DEGF$

Using the polygon area formula,

$$[DEGF] = \frac{1}{2}\left|\vec{D}\times\vec{E} + \vec{E}\times\vec{G} + \vec{G}\times\vec{F} + \vec{F}\times\vec{D}\right|.$$

Compute each term:

$$\vec{D}\times\vec{E} = 0, \quad \vec{G}\times\vec{F} = 0,$$

$$\vec{E}\times\vec{G} = \frac{10}{21}(\vec{b}\times\vec{c}), \quad \vec{F}\times\vec{D} = -\frac{2}{105}(\vec{b}\times\vec{c}).$$

Thus,

$$[DEGF] = \frac{1}{2}\cdot\frac{16}{35}|\vec{b}\times\vec{c}| = \frac{8}{35}\cdot 2730\sin\theta = 624\sin\theta.$$

Given

$$[DEGF] = 288,$$

we obtain

$$\sin\theta = \frac{288}{624} = \frac{6}{13}.$$

Step 6: Area of triangle ABC

$$[\triangle ABC] = \frac{1}{2}\cdot 30\cdot 91\cdot\sin\theta = 1365\cdot\frac{6}{13} = 630.$$

Conclusion

The remaining task is to express the area of heptagon $AFNBCEM$ in terms of the area of triangle ABC, using reflections about points F and E.

A.3 Phase 1: Truncation Points

We detect the potential collapse and truncate the erroneous trajectory at three distinct milestones (t_1, t_2, t_3).

- **Truncation Point 1 (t_1):**

  We are given:

  - Points D, E on AB such that $AD < AE < AB$,
  - Points F, G on AC such that $AF < AG < AC$,
  - Lengths:
    - $AD = 4$
    - $DE = 16$
    - $EB = 8$
    - $AF = 13$
    - $FG = 52$
    - $GC = 26$
  - M is the reflection of D through F,
  - N is the reflection of G through E,
  - Area of quadrilateral $DEGF = 288$,
  - We are to find the area of heptagon $AFNBCEM$.

Step 1:

- **Truncation Point 2 (t_2):**

  We are given:

  - Points D, E on AB such that $AD < AE < AB$,
  - Points F, G on AC such that $AF < AG < AC$,
  - Lengths:
    - $AD = 4$
    - $DE = 16$
    - $EB = 8$
    - $AF = 13$
    - $FG = 52$
    - $GC = 26$
  - M is the reflection of D through F,
  - N is the reflection of G through E,
  - Area of quadrilateral $DEGF = 288$,
  - We are to find the area of heptagon $AFNBCEM$.

  Step 1: Understand the given structure

  On AB:

  $$AD = 4, \quad DE = 16, \quad EB = 8$$

  $$AB = AD + DE + EB = 4 + 16 + 8 = 30$$

  On AC:

  $$AF = 13, \quad FG = 52, \quad GC = 26$$

  $$AC = AF + FG + GC = 13 + 52 + 26 = 91$$

  Step 2:

- **Truncation Point 3 (t_3):**

  We are given:

  - Points D, E on AB such that $AD < AE < AB$,
  - Points F, G on AC such that $AF < AG < AC$,
  - Lengths:
    - $AD = 4$
    - $DE = 16$
    - $EB = 8$
    - $AF = 13$
    - $FG = 52$
    - $GC = 26$
  - M is the reflection of D through F,
  - N is the reflection of G through E,
  - Area of quadrilateral $DEGF = 288$,
  - We are to find the area of heptagon $AFNBCEM$.

  Step 1: Understand the given structure

  On AB:

  $$AB = 4 + 16 + 8 = 30$$

  On AC:

  $$AC = 13 + 52 + 26 = 91$$

  Step 2: Understand points M and N

  - M is the reflection of D through F, so F is the midpoint of DM.
  - N is the reflection of G through E, so E is the midpoint of GN.

  Since D lies on AB and F lies on AC, the reflection is a point reflection in the plane, not along a line.

  We are asked to compute the area of the heptagon:

  $$AFNBCEM$$

  Step 3:

A.4 Phase 2: Micro-SVD & Probe Selection (At Truncation Point t_1)

Here we detail the geometric intervention at the first truncation point.

A.4.1 Manifold Estimation (Monte Carlo Samples)

Taking truncation point 1 as an example, the Teacher generates N look-ahead paths to estimate the local covariance matrix Σ_{t_1}. Note how these samples are semantically similar (confirming the Bias Manifold).

MC Sample 1: Understand the configuration. We have triangle

MC Sample 2: Understand the configuration. We have triangle

MC Sample 3: Understand the configuration. We have triangle

MC Sample 4: Understand and use given lengths. We

MC Sample 5: Understand the configuration. We are dealing with

MC Sample 6: Understand the configuration. We have triangle

MC Sample 7: Understand the configuration. We have triangle

MC Sample 8: Understand the configuration. Let's first sketch

A.4.2 Student Probe Candidates ($L = 8$ tokens)

The Student model (π_ϕ) generates 8 candidate sequences. We calculate the **Orthogonality Score** for each relative to the Teacher's estimated manifold.

MC Sample 1: Find AE, AB, AC. Orthogonality Score: 0.9403

MC Sample 2: Let $A(x_1$. Orthogonality Score: 0.9844

MC Sample 3: Let $A = (0$. Orthogonality Score: 0.9802

MC Sample 4: First, let's find the Orthogonality Score: 0.7196

MC Sample 5: Let $A = (0,$. Orthogonality Score: 0.9801

MC Sample 6: We are given the lengths AD. Orthogonality Score: 0.8500

MC Sample 7: We know $AB = AD+$. Orthogonality Score: 0.9643

MC Sample 8: **Since $AD = 4$, ... Orthogonality Score: 0.9724**

B Reasoning Collapse Evidence

To verify the universality of the observed geometric pathology, we extended our spectral analysis to models of varying scales, specifically `Qwen3-8B`(Team, 2025b) and `QwQ-32B`(Team, 2025c). Employing the identical methodology on the AIME 2025 benchmark, we tracked the effective rank dynamics of the hidden state covariance matrices during inference. As illustrated in Figure 7 and Figure 8, both models exhibit a consistent pattern of spectral degeneration: despite their increased parameter counts, the effective rank decays significantly as the reasoning chain lengthens and enters erroneous loops. This evidence suggests that reasoning collapse is a fundamental geometric phenomenon that persists across model scales, rather than an artifact specific to smaller architectures.

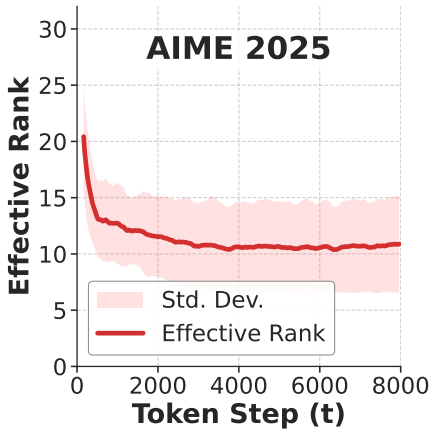


Figure 7: Reasoning Collapse of 8B Model

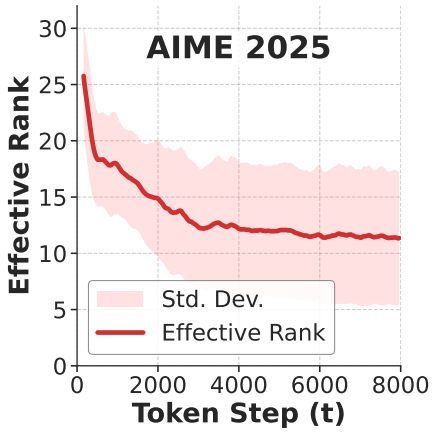


Figure 8: Reasoning Collapse of 32B Model

C Ablation Studies and Robustness Analysis

C.1 Ablation Studies

To rigorously evaluate the contribution of each component in the Spectral Orthogonal Exploration (SOE) framework, we conduct ablation studies focusing on the necessity of the probe injection and the effectiveness of the geometric selection mechanism. Due to computational resource constraints, these analyses were performed on the AIME 2025 benchmark.

Validity of the Baseline as an Ablation. We first emphasize that the **Baseline (Self-Consistency)** reported in the main text serves as the fundamental ablation of our framework. It represents the "No Injection" setting, where the Teacher model samples repeatedly from the same collapsed trajectory without any external intervention. The significant performance gap between the Baseline and SOE (as shown in the main results) confirms the essential value of the "Student Guides Teacher" paradigm.

Effectiveness of Micro-SVD Selection (Random vs. Orthogonal). To isolate the impact of our *Orthogonal Latent Stitching* (Micro-SVD) mechanism, we conducted a further ablation. In this setting, the Student generates candidate probes, but instead of selecting the candidate with the highest orthogonality score, we select one **at random** to stitch into the Teacher's context.

This comparison allows us to determine whether performance gains arise merely from adding "noise" (random perturbation) or from the specific geometric directionality of our method. As shown in Table 8, while random injection provides a performance boost over the baseline (58.82% vs 35.29%) by disrupting the bias manifold, our proposed SOE method achieves a significantly higher accuracy (70.59%). This result confirms that the Micro-SVD mechanism effectively identifies high-value directions in the orthogonal complement that are superior to random perturbations.

Sampling Efficiency. Figure 9 further illustrates the sampling efficiency comparison. Our method reaches high-confidence solutions with fewer samples compared to both the Baseline and the Random Ablation, indicating that SOE navigates the solution space more efficiently.

Method	Pass@16 (AIME 25)
Baseline	35.29%
Ablation	58.82%
Ours (SOE)	**70.59%**

Table 8: Ablation study on AIME 2025. "Random Probe" indicates selecting a Student trace without geometric filtering. The results demonstrate that while any external diversity helps, directed orthogonal exploration yields the optimal performance.

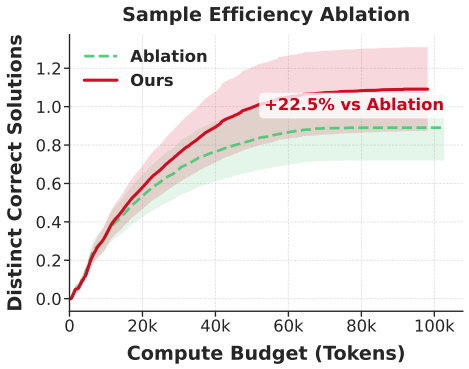


Figure 9: Sampling efficiency comparison between Baseline, Random Ablation, and SOE on AIME 2025. SOE discovers correct solutions at a significantly faster rate.

C.2 Additional Analysis of Exploration Efficiency

Orthogonality Score Distribution. To validate the geometric mechanism underlying SOE, we analyze the distribution of Orthogonality Scores for Student-generated probes across benchmarks.

As shown in Figure 10, the Student-generated probes consistently exhibit high orthogonality relative to the Teacher's local Bias Manifold. This supports our claim that the Student provides sufficiently heterogeneous reasoning signals to project into the Teacher's orthogonal complement, thereby helping escape low-rank confinement.

C.3 Sensitivity Analysis of the Diversity Threshold

To test the robustness of the *Distinct Correct Solutions* metric, we vary the cosine-similarity threshold τ used to distinguish semantically different correct solutions. Table 9 reports the results on AIME 2024 and AIME 2025. Across all threshold choices, SOE consistently discovers more distinct correct solutions than the baseline, confirming that the improvement in exploration efficiency is robust to the specific choice of τ.

Cosine Threshold	AIME 25 (Baseline)	AIME 25 (Ours)	AIME 24 (Baseline)	AIME 24 (Ours)
0.80 (Strict)	0.244	**0.886**	0.265	**0.944**
0.90	0.356	**1.660**	0.495	**1.971**
0.95 (Original)	0.640	**2.482**	1.028	**3.258**
1.00 (Exact Match)	0.740	**3.080**	1.028	**4.285**

Table 9: Sensitivity analysis of the diversity threshold τ used in the *Distinct Correct Solutions* metric. SOE consistently finds more distinct correct solutions than the baseline across all tested thresholds.

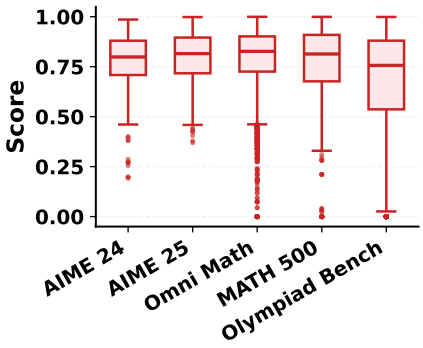


Figure 10: **Distribution of Orthogonality Scores** for Student-generated probes across benchmarks. The score ranges from 0 to 1, where a value of 1 indicates perfect orthogonality relative to the Teacher's Bias Manifold, and 0 denotes complete alignment (collinearity).

C.4 Robustness Analysis

To evaluate the stability of our method against stochastic variations, we conducted repeated experiments on the AIME 2025 benchmark using three distinct random seeds ($\{42, 1024, 2024\}$). Due to computational resource constraints, this robustness analysis was restricted to the AIME 2025 dataset. As shown in Table 10, our method (SOE) consistently outperforms the baseline across all random seeds, demonstrating that the performance improvements are not artifacts of a specific initialization. Notably, while the baseline fluctuates between 29.41% and 41.18%, our method achieves a stable accuracy of up to 70.59%.

Seed	Baseline	Ours (SOE)
42	29.41%	64.71%
1024	41.18%	70.59%
2024	35.29%	70.59%

Table 10: Robustness analysis on AIME 2025 across three different random seeds. Our method demonstrates consistent performance gains despite stochastic variations.

D Theoretical Proofs and Extended Analysis

D.1 Proof of Proposition 1 and Analysis of Non-Linearity

Proof. Consider a simplified linear attention head. The output for a query q is given by:

$$\text{Attn}(q, K, V) = \sum_{i=1}^{N} \text{softmax}(q^T k_i) v_i \quad (10)$$

where $v_i = W_V h_i$. Since each input $h_i \in \mathcal{S}$, each value vector v_i lies in $W_V \mathcal{S}$. The output is therefore a convex combination of vectors in $W_V \mathcal{S}$. Thus, the generated token embedding remains within the span of the existing history (up to the corresponding linear transform). Without an external perturbation or a mechanism that introduces new directions, the model has limited ability to explore the orthogonal complement $\mathcal{S}^{\perp}$. □

Remark 1: The Limited Role of Pointwise Non-Linearity. A natural counter-argument is that the Feed-Forward Networks (FFNs) in Transformers, which contain non-linear activation functions (such as GELU or ReLU), might restore lost dimensionality. We analyze this possibility by examining the coordinate-wise nature of these activations.

Consider the FFN update: $h' = W_2 \sigma(W_1 h)$. The linear transformations W_1, W_2 satisfy $\text{rank}(Wh) \leq \text{rank}(h)$, so any dimensional expansion would need to arise from the non-linearity $\sigma(\cdot)$.

However, standard activation functions are **coordinate-wise**: $\sigma(v)_i = \sigma(v_i)$. They operate on each dimension independently without mixing information.

- **Zero-Variance Persistence**: If the input distribution has collapsed such that variance is zero along a basis vector e_j, then for any sample h, the j-th component is constant (or zero). Since σ acts independently, the output $\sigma(h_j)$ remains constant. The activation does not by itself introduce variation in that direction.

- **Manifold Curvature vs. Expansion**: While non-linearity can bend a flat k-dimensional subspace into a curved k-dimensional manifold, it does not necessarily increase the *intrinsic dimensionality*. A piece of paper crumpled into a ball still has intrinsic dimension 2; it does not acquire volume.

Consequently, without an external geometric intervention, the FFN may warp the collapsed Bias Manifold but does not necessarily restore exploration in previously inactive directions (Geva et al., 2021).

D.2 Proof of Proposition 2

Proof. Let the attention weights be $A \in \mathbb{R}^{N\times N}$, and consider the simplified update rule

$$H_{t+1} = AH_t. \tag{11}$$

In later stages of reasoning, the model may exhibit a form of confirmation bias in which many tokens attend heavily to a small set of anchor tokens. In this regime, the rows of the attention matrix can become increasingly similar, reducing the diversity of directions that contribute to the next hidden-state update.

Using the singular-value inequality $\sigma_i(XY) \leq \|X\|\sigma_i(Y)$, we obtain that

$$\operatorname{rank}(H_{t+1}) \leq \min\big(\operatorname{rank}(A), \operatorname{rank}(H_t)\big). \tag{12}$$

Therefore, the attention update satisfies a non-expansion-style constraint on algebraic rank. When combined with increasingly concentrated attention patterns, this provides a theoretical perspective on why the representation may become progressively more spectrally concentrated over time. □

D.3 Semantic Sparsity Analysis

Here we discuss the informational contribution of the injection. Let the injected Student sequence be s_{inj} with length L_{inj} (8 tokens in our experiments), and let the Teacher's context length be L_{ctx} (8192 tokens in our experiments).

Argument: Semantic Sparsity. The direct contribution of the injection to the global context mean is small:

$$\Delta\mu \approx \frac{L_{\text{inj}}}{L_{\text{ctx}}}(h_{\text{Student}} - \mu_{\text{Teacher}}) \approx 0 \tag{13}$$

Since $L_{\text{inj}} \ll L_{\text{ctx}}$, the injected sequence contributes only a limited shift to the global centroid of the context. This observation is consistent with the view that SOE's gains are not solely explained by large-scale information transfer through the added tokens. Instead, the injected probe may be more effective as a localized geometric perturbation that changes subsequent attention routing and exploration.

E LLM Usage Statement

Large Language Models (LLMs) such as ChatGPT and Gemini were used as general-purpose writing assistants to improve readability and clarity of the manuscript (*e.g.*, grammar checking, LaTeX formatting, and sentence restructuring). LLMs were not used to determine the research idea, perform literature search or generate the paper's key insights and conclusions. All analyses and conclusions are solely those of the authors.